INFORMATION AND CONTROL 37, 90–99 (1978)

Optimal Control-Limit Policies for a Zero-Memory Replacement Problem*

CHELSEA C. WHITE III

Department of Applied Mathematics and Computer Science, University of Virginia, Charlottesville, Virginia 22903

In this paper, we investigate the structural properties of optimal policies for a finite-memory, stochastic replacement problem. The characteristic which distinguishes this problem from the usual replacement problem is its information pattern; it is assumed that the controller has available for decision making the present (but not any past) realization of a process stochastically related to the state process. Two classes of structured policies are defined, both of which are generalizations of the class of control-limit policies for the usual replacement problem. For each definition, we present conditions which are proved to be sufficient for the existence of optimal control-limit policies.

1. INTRODUCTION

The investigation of explicit structural properties of optimal control laws (and their computational implications) for special multistage decision problems has been a primary area of research in the control literature. The linear–quadratic–Gaussian (LQG) problem (see, for example, Kushner, 1971) serves as a prominent example for continuous state stochastic control problems; stochastic inventory models (Scarf, 1963) and replacement models (Derman, 1963) are well-known examples for countable state control problems.

With regard to the LQG problem, there has been considerable interest in exploring the relationship between the functional form of the optimal control law and the problem's *information pattern*, i.e., what the controller knows and when the controller knows it. It is well known that if the information pattern is *classical* (if all controllers have perfect memory and totally share both their data and structural information), there exists under mild assumptions an optimal control law which is linear in the state vector (Kushner, 1971). This explicit functional form of the optimal control law is retained for certain nonclassical, delayed data sharing information patterns (Kurtaran, 1975) but does not hold for other, finite-memory information patterns (Witsenhausen, 1968).

* This research has been supported by NSF Grants ENG75-19692 and ENG76-15774.

0019-9958/78/0371-0090$02.00/0

The study of explicit structural form of optimal policies (and most other studies as well) for countable state stochastic control problems has dealt almost exclusively with problems where the controller has perfect memory and perfect access to the state of the system. We will refer to this special case of the classical information pattern as the *standard* information pattern. Of times, however, the standard information pattern is not a realistic model of the controller's state of knowledge. As examples, a machine inspector may not know the actual state of the machine but may have available the items produced by the machine; a physician may not know the underlying state of the patient's health but may have available various signs, symptoms, and laboratory test results. If the machine inspector and the physician keep formerly obtained data, i.e., they essentially have perfect memory, then an adequate model of what they know and when they know it is the classical information pattern, where decisions are based on realizations of processes that are statistically related to the state process. Other examples of systems subject to partial observation are discussed by Smallwood and Sondik (1973) and White (1976a). If the perfect memory assumption is not valid and/or if there are several controllers who do not (or cannot) share their structural information and on-line data, then the *finite-memory* information pattern (Sandell, 1974) may be a realistic model of the controllers' state(s) of knowledge. This paper presents a replacement problem with a zero-memory information pattern. Other examples, such as a data communications problem, are presented by Sandell (1974).

Knowledge of the structural properties of optimal policies for a problem having a nonstandard information pattern has implied substantial computational simplification (White, 1976b). Similar computational simplifications may also result from optimal policy structural properties for other problems having nonstandard information patterns. Such simplification would be particularly desirable since the usefulness of problems with nonstandard information patterns if often severely limited by computational complexity. (See Sandell, 1974, and Sondik, 1971, for discussions of the computational requirements for countable state control problems having nonstandard information patterns.)

In this paper, we investigate the structural properties of optimal policies for a special dynamic programming problem having a nonstandard information pattern. Specifically, a replacement problem is studied where the controller has only the present noise corrupted observation of the state process on which to base a decision. We show that under reasonable assumptions, there exist optimal policies having a functional form which is a generalization of a control-limit policy. The assumptions made generalize but closely resemble sufficient conditions for the existence of optimal control-limit policies for the standard information pattern case (Derman, 1963). Two generalized definitions of the usual control-limit policy are considered, and conditions which guarantee that we only need to examine policies with these structural attributes are presented for both definitions.

The paper is organized as follows. Section 2 presents the finite-horizon and discounted cost zero-memory replacement problems and several equivalent problem formulations. The original problem formulations are first recast into the finite-state, finite-memory problem description presented by Sandell (1974). We then present the equivalent deterministic control problem (as described and analyzed by Sandell, 1974) and formulate it as a special case of the usual Markov decision problem with standard information pattern, using a notation similar to that which is found by Porteus (1975).

Using results of White (1975), we begin Section 3 by reducing the examination of the structural form of optimal policy for the zero-memory problem to the examination of the optimal policy structure for a special case of the partially observed problem studied by White (1976c). This reduction leads directly to the first set of conditions sufficient for the existence of optimal control-limit policies. An additional assumption is then shown to imply that there exist optimal policies having a further restricted control-limit form.

2. Problem Formulation

The zero-memory replacement problem is now defined. Let the discrete time stochastic process $\{s(t), t = 0, 1,...\}$, having finite-state space S, model a system subject to replacement and Markov deterioration. Assume $\leqslant_S$ is a partial ordering on S which describes the relative level of deterioration between two states, i.e., $i \leqslant_S i'$ if and only if state i is at least as "good" as state i'. Let $O \in S$ designate the state of a new system, and hence $O \leqslant_S i$ for all $i \in S$. We also assume that the a priori probabilities $\xi = \{\xi_i\}$ are given, where $\xi_i = P[s(0) = i]$.

The state of the system is observed at each time $t = 0, 1,...,$ where the observation at time t is the realization of the random variable $z(t)$, which is assumed to have finite-state space Z.

Let $n \leqslant \infty$ represent the terminal time of the control problem. At each time $t = 0,..., n - 1$, the controller is allowed to choose either to replace the system or to do nothing. A replace decision sends the state of the system to O just prior to the next decision epoch; a do nothing decision allows the system to be subject to Markov deterioration over the next time interval. The decision made at time t is represented by $u(t) \in C = \{0, 1\}$, where $0 =$ do nothing and $1 =$ replace. The controller is assumed to base his (her or its) decision on only the present realization of the observation random variable; that is,

$$u(t) = y_t[z(t)]. \tag{1}$$

This zero-memory assumption represents the key aspect which distinguishes this problem formulation from other replacement problems. Equation (1) states that the controller "forgets" all former realizations of the process $\{z(t), t = 0, 1,...\}$. Such an information pattern may be realistic for systems with limited memory

computer-based controllers or serve as an approximation to human decision makers with less than perfect recall and/or system record availability.

The above processes are related by the conditional probabilities $p_{ij}(c) = P[s(t+1) = j \mid s(t) = i,\ u(t) = c]$ and $q_{jk} = P[z(t) = k \mid s(t) = j]$, where $p_{ij}(0) = p_{ij}$, $p_{ij}(1) = 1$ if $j = O$, $P(c) = \{p_{ij}(c)\}$, and $Q = \{q_{jk}\}$.

The cost structure is defined as follows. Let $g[s(t), u(t)] \geqslant 0$ be the cost accrued at time $t < n$. When the problem horizon is finite, i.e., when $n < \infty$, a terminal cost $g_0[s(n)] \geqslant 0$ is additionally accrued at the terminal time. The finite horizon problem is to select a sequence of controls y_t, $t = 0, 1,..., n-1$, satisfying (1) which minimizes the criterion $E\{\sum_{t=0}^{n-1} \beta^t g[s(t), u(t)] + \beta^n g_0[s(n)] \mid \xi\}$, where $\beta \in [0, 1]$ is the discount factor. Likewise, the discounted cost problem is to select a sequence of controls y_t, $t = 0, 1,...,$ satisfying (1) which minimizes $E\{\sum_{t=0}^{\infty} \beta^t g[s(t), u(t)] \mid \xi\}$, where $\beta \in [0, 1)$.

These problems are examples of the finite-state, finite-memory stochastic control problems that have been formulated and analyzed by Sandell (1974). We now restate the finite-memory replacement problems as special cases of the problems considered by Sandell (1974). The state space of the restated problems is $\mathscr{S} = S \times Z$. The stochastic process which we wish to control is $\{\sigma(t),\ t = 0, 1,...\}$, where $\sigma(t) = \{s(t), z(t)\}$. The new state space therefore is ordered by the relation $\leqslant$, where for $i, i' \in \mathscr{S}$, $i = (i_1, i_2)$ and $i' = (i_1', i_2')$, $i \leqslant i'$ if and only if $i_1 \leqslant_S i_1'$. The new state process evolves according to the transition probabilities $\{\not{p}_{ij}(y)\}$, where $\not{p}_{ij}(y) = p_{i_1 i_2}[y(i)] q_{j_1 i_2}$. Controls are of the form $u(t) = y_t[\sigma(t)]$, where $y_t \in D = \{y : y \text{ maps } \mathscr{S} \text{ into } C \text{ and } y(i_1, i_2) = y(i_1', i_2)\}$. Note that D is equivalent to $C^{\mathscr{N}}$, where $\mathscr{N} = \text{card}(Z)$. The cost accrued at time $t < n$ is $\not{g}[\sigma(t), y_t] = g[s(t), u(t)]$, where $u(t) = y_t[\sigma(t)]$; the terminal cost accrued is $\not{g}_0[\sigma(n)] = g_0[s(n)]$. For the restated version of the finite-horizon problem, we wish to select a sequence of controls $y_t \in D$, $t = 0,..., n-1$, which minimizes the criterion $E\{\sum_{t=0}^{n-1} \beta^t \not{g}[\sigma(t), y_t] + \beta^n \not{g}_0[\sigma(n)] \mid \xi\}$; the discounted cost case is defined accordingly.

It has been shown by Sandell (1974) that the above finite-state, finite-memory stochastic control problems have equivalent deterministic control problem formulations. These formulations will be of primary interest throughout the remainder of this paper and are described as follows. The state space of the deterministic problems is $\Omega = \{x : x_i \geqslant 0 \text{ and } \sum_{i \in \mathscr{S}} x_i = 1\}$, which is the set of probability vectors on $\mathscr{S}$. The state of the deterministic problems evolves according to the difference equation $x(t+1) = x(t)\, \mathscr{P}(y_t)$, where the jth element of $x\mathscr{P}(y)$, $j \in \mathscr{S}$, is $\sum_{i \in \mathscr{S}} x_i \not{p}_{ij}(y)$ and where we choose $x_i(0) = \xi_{i_1}/\mathscr{N}$ for $i = (i_1, i_2)$. At time t, $0 \leqslant t < n$, $r(x, y) = \sum_{i \in \mathscr{S}} x_i \not{g}[i, y(i)]$ is accrued when $x(t) = x$ and $y_t = y$; similarly, the terminal cost $r_0(x) = \sum_{i \in \mathscr{S}} x_i \not{g}_0(i)$ is accrued when $x(n) = x$ at the terminal time for the finite horizon problem. The equivalent finite horizon deterministic control problem is to select a sequence of controls $y_t \in D$, $t = 0,..., n-1$, which minimizes $\sum_{t=0}^{n-1} \beta^t r[x(t), y_t] + \beta^n r_0[x(n)]$; the discounted cost problem is similarly defined.

Since $x(0)$ and $\mathscr{P}(y)$ are known to the controller in the deterministic problems the sequence of controls can be chosen as a function of $x(t)$, $t = 0, 1, \ldots$. Thus, the deterministic control problems can be equivalently stated as follows: select of sequence of control policies $\{\delta_0, \delta_1, \ldots\}$ which minimizes the appropriate criterion, where each policy is a member of the *policy space* $\Delta = D^{\Omega}$; i.e., Δ is the set of all functions of the form $\delta: \Omega \to D$.

3. Optimal Policy Structure

The intent of this section is to present conditions which imply that only strict subsets of the policy space Δ need be examined in searching for optimal policies. Two subsets of Δ are considered, both representing different interpretations of the set of all control-limit policies. Our developments require several preliminary definitions, most of which will be familiar to readers of Porteus (1975) and his references.

The optimal expected cost to be accrued between time t and the terminal time n is $f_{n-t} = \inf(H_{\delta_t} \times \cdots \times H_{\delta_{n-1}} r_0)$, the infimum is with respect to the set of all admissible *strategies*, i.e., sequences $(\delta_t, \ldots, \delta_{n-1})$ in Δ^{n-t}, $[H_\delta v](x) = h[x, \delta(x), v]$, and $h(x, y, v) = r(x, y) + \beta v[x\mathscr{P}(y)]$. Similarly, the optimal expected discounted cost accrued over the infinite horizon is

$$f = \lim_{n\to\infty} \inf(H_{\delta_t} \times \cdots \times H_{\delta_{n-1}} r_0).$$

Define the operator A as $Av = \inf_{\delta\in\Delta} H_\delta v$. It is well known that $\{f_n\}$ can be described recursively by the dynamic programming equation $f_n = Af_{n-1}$ and that for $\beta < 1$, f is the fixed point of the contraction A (Denardo, 1967). (The complete metric space on which A is a contraction is defined by Denardo, 1967.)

Since the $z(t)$ element of the state vector $\mathscr{s}(t)$ is completely observed, the optimal control strategies have a simplified form, which will be presented after further preliminary definitions. Let $\pi^k(x) = \sum_i x_{ik}$, where $x_{ik} = P[s(t) = i, z(t) = k]$. Thus, $\pi^k(x) = P[z(t) = k]$, $\pi^k: \Omega \to [0, 1]$ and $\sum_k \pi^k(x) = 1$. Define $\Xi = \{\eta : \eta_i \geqslant 0, \sum_{i\in S} \eta_i = 1\}$. Let $\Pi^k: \Omega \to \Xi$ have ith component $\Pi_i^k(x) = x_{ik}/\pi^k(x)$ if $\pi^k(x) \neq 0$. (When $\pi^k(x) = 0$, define $\Pi^k(x)$ in Ξ arbitrarily.) Note that $\Pi_i^k(x) = P[s(t) = i \mid z(t) = k]$; hence, $\Pi^k(x) \in \Xi$ is the a posteriori density vector for $s(t)$ if the a priori density matrix for the pair $\{s(t), z(t)\}$ is $x \in \Omega$ and if the realization of $z(t)$ is known to be k (which occurs with probability $\pi^k(x)$).

Define $\tilde{h}(\eta, c, v) = \sum_i g(i, c) + \beta \sum_k \sigma(k, \eta, c)\, v[\lambda(k, \eta, c)]$ and $[\tilde{H}_{\tilde{\delta}} v](\eta) = \tilde{h}[\eta, \tilde{\delta}(\eta), v]$, where $\tilde{\delta} \in \tilde{\Delta} = C^{\Xi}$, $v \in R^{\Xi}$, $\sigma(k, \eta, c) = \sum_{i\in S} \sum_{j\in S} q_{jk} p_{ij}(c) \eta_i$, $\lambda_j(k, \eta, c) = q_{jk} \sum_i p_{ij}(c)\, \eta_i / \sigma(k, \eta, c)$, and $\lambda(k, \eta, c) = \{\lambda_j(k, \eta, c)\} \in \Xi$. (If $\sigma(k, \eta, c) = 0$, define $\lambda(k, \eta, c)$ arbitrarily in Ξ.) The functions σ and λ have the following interpretation. At time t, $x(t)$ and $z(t)$ are known, and hence it is sufficient to know only $\Pi^{z(t)}[x(t)]$ for a probabilistic description of the realization

of $s(t)$. Bayes' rule shows that $P[z(t+1) = k \mid x(t) = x, z(t) = l, u(t) = c] = P[z(t+1) = k \mid \Pi^l(x) = \eta, u(t) = c] = \sigma(k, \eta, c)$ and that $P[s(t+1) = j \mid z(t+1) = k, x(t) = x, z(t) = l, u(t) = c] = P[s(t+1) = j \mid z(t+1) = k, \Pi^l(x) = \eta, u(t) = c] = \lambda_j(k, \eta, c)$.

Let the operator $\tilde{A}$ be such that $\tilde{A}v = \inf_{\tilde{\delta}} \tilde{H}_{\tilde{\delta}} v$. Define $\tilde{f}_n$ by $\tilde{f}_n = \tilde{A}\tilde{f}_{n-1}$, $\tilde{f}_0 = \sum_i \eta_i g_0(i)$, and for $\beta < 1$ let $\tilde{f}$ be the fixed point of $\tilde{A}$. The function $\tilde{f}_n \in R^{\Xi}$ is the optimal expected cost to be accrued over a finite horizon of length n and has as its argument a probability density vector on S (which is a member of Ξ). The complete observability of $z(t)$ has been shown by White (1975) (generalizing results due to Astrom, 1965) to imply that f_n and $\tilde{f}_n$ are related by the equation $f_n(x) = \sum_k \pi^k(x) \tilde{f}_n[\Pi^k(x)]$ for all n and hence for $\beta < 1$ $f(x) = \sum_k \pi^k(x) \tilde{f}[\Pi^k(x)]$. Let $\tilde{\delta}_t \in \tilde{\Delta}$ ($\delta_t \in \Delta$) be an optimal policy which achieves the infimum of $\tilde{A}\tilde{f}_{n-t-1}$ (Af_{n-t-1}). It has been shown by White (1975) that δ_t can be constructed as follows: if $z(t) = k$, then choose $\delta_t(x) = \tilde{\delta}_t[\Pi^k(x)]$. That is, the optimal policy depends only on the conditional density of the state given the present observation. Thus, the problem of determining δ_t and f_{n-t}, both depending on an argument having card ($\mathscr{S}$) elements, can be reduced to the problem of determining $\tilde{\delta}_t$ and $\tilde{f}_{n-t}$, which both depend on an argument having card (S) ($\leqslant$ card ($\mathscr{S}$)) elements. This result allows us to restrict attention to the subset $\bar{\Delta} \subset \Delta$ with the following property: for all $x \in \Omega$ there exists a $\tilde{\delta} \in \tilde{\Delta}$ such that $\delta(x) = \tilde{\delta}[\Pi^k(x)]$ if the present realization of the observation process is k.

We have thus far reduced the original problem of selecting a $\delta \in \Delta$ to minimize $H_\delta f_{n-t-1}$ to a problem of selecting a $\tilde{\delta} \in \tilde{\Delta}$ to minimize $\tilde{H}_{\tilde{\delta}} \tilde{f}_{n-t-1}$. We now define structural properties on $\bar{\Delta}$ in terms of structural properties on $\tilde{\Delta}$.

DEFINITION 1. (a) Let the partial ordering $\prec$ on Ξ be defined as $\eta \prec \eta'$ if and only if $\eta I_K \leqslant \eta' I_K$ for all $K \in \mathscr{K} = \{K \subset S : i \in K \text{ and } i \leqslant_S i' \text{ implies } i' \in K\}$, where I_K is the indicator function of the set K and where $\eta I_K = \sum_{i \in S} \eta_i I_K(i)$.

(b) Let the partial ordering $\ll$ on Ω be defined as $x \ll x'$ if and only if $\Pi^k(x) \prec \Pi^k(x')$ for all $k \in Z$.

We say $\delta(x) \leqslant \delta(x')$ for $\delta \in \bar{\Delta}$ if and only if $\tilde{\delta}[\Pi^k(x)] \leqslant \tilde{\delta}[\Pi^k(x')]$ for all $k \in Z$, where $\tilde{\delta}$ is the element in $\tilde{\Delta}$ associated with $\delta \in \bar{\Delta}$. The notion of a control-limit policy is now extended to the case where the state space is a probability vector.

DEFINITION 2. (a) The set of all control-limit policies on $\tilde{\Delta}$ is $\tilde{\Delta}' = \{\tilde{\delta} \in \tilde{\Delta} : \eta \prec \eta' \text{ implies } \tilde{\delta}(\eta) \leqslant \tilde{\delta}(\eta')\}$.

(b) The set of all control-limit policies on $\bar{\Delta}$ is $\Delta' = \{\delta \in \bar{\Delta} : x \ll x' \text{ implies } \delta(x) \leqslant \delta(x')\}$.

It is easily shown that if $z(t) = s(t)$ with probability one for all t, then the above definitions are equivalent to the usual control-limit policy definition for the completely observed case presented by Derman (1963).

Our main result will be to show that if three assumptions are satisfied, attention can be restricted to Δ' in selecting optimal policies. These assumptions are stated following further preliminary definitions.

Let $\mathscr{F}$ be the set of all functions α: $S \to R$ (where R is the real line) which are nondecreasing with respect to $\leqslant_S$, i.e., $\alpha \in \mathscr{F}$ if and only if $i \leqslant_S i'$ implies $\alpha(i) \leqslant \alpha(i')$. Let the (j, j)th element of the array R_k be q_{jk} with all other elements of R_k set to zero. (When both S and Z are linearly ordered, $R_k = \text{diag}\{q_{jk}\}$, a diagonal matrix.) Define the sets $\mathscr{A}_n$ as: $\mathscr{A}_0 = \{g_0\}$ $\mathscr{A}_n = \{g(c) + \beta P(c) \sum_k R_k \alpha^k$: $c \in C$ and $\alpha^k \in \mathscr{A}_{n-1}\}$, where the ith element of $\mathscr{A}_n$, $n > 0$, is $g(i, c) + \beta \sum_j p_{ij}(c) \sum_k q_{jk} \alpha_k{}^j$, for $\alpha^k \in A_{n-1}$.

ASSUMPTIONS. (A1) $g(\cdot, c), g_0(\cdot)$, and $g(\cdot, 0) - g(\cdot, 1)$ are all members of $\mathscr{F}$ for $c \in C$.

(A2) $PI_K \in \mathscr{F}$ for all $K \in \mathscr{K}$, where the ith element of PI_K is $\sum_j p_{ij} I_K(j)$.

(A3) For each $t \in \{0,..., n\}$, $\sum_k R_k \alpha^k \in \mathscr{F}$, where $\alpha^k \in \mathscr{A}_t$ for all k.

Assumptions (A1) and (A2) are slight generalizations of the usual sufficient conditions for the existence of optimal control-limit policies for the completely observed case. (A1) stipulates that increased cost must be consistent with the notion of increased system deterioration; (A2) is the familiar increasing failure rate assumption due to Derman (1963). Both of these assumptions have been shown to imply the existence of optimal control-limit policies for the completely unobserved information pattern (White, 1976b). A counterexample presented by White (1976b) also showed that in general (A1) and (A2) are not enough to insure the existence of optimal control-limit policies for the general partially observed case. It was proved in (White, 1976c) that the inclusion of (A3) with (A1) and (A2) does imply the desired existence result; conditions on the problem formulation which imply (A3) to hold are also presented in (White, 1976c). Thus, these results state that (A1), (A2), and (A3) are sufficient conditions for the existence of an optimal control-limit policy for the control problem associated with $\tilde{f}_n$ and $\tilde{f}$; that is, for each $n = 1, 2,...$, and for $n = \infty$ if $\beta < 1$, there exists a $\tilde{\delta} \in \tilde{\Delta}'$ such that $\tilde{f}_n = \tilde{H}_{\tilde{\delta}} \tilde{f}_{n-1}$ and $\tilde{f} = \tilde{H}_{\tilde{\delta}} \tilde{f}$. We now can state our main result.

THEOREM. *Assume* (A1), (A2), *and* (A3). *Then, for the finite horizon case there exists an optimal strategy composed of policies that are members of* Δ'. *For the discounted cost case* ($\beta < 1$), *there exists an optimal stationary strategy generated by a policy in* Δ'.

Proof. The comments preceding the theorem statement guarantee the existence of an optimal policy in $\tilde{\Delta}'$ for the control problem associated with $\tilde{f}_n$ and $\tilde{f}$. This optimal policy in $\tilde{\Delta}$ has a related policy in Δ which is necessarily optimal for the control problem associated with f_n and f. Let $\tilde{\delta} \in \tilde{\Delta}'$ and $\delta \in \Delta$

be these policies; thus, $\delta(x) = \{\tilde{\delta}[\Pi^k(x)]\}$. The definitions then directly imply that $\delta \in \Delta'$. The existence of an optimal stationary policy is due to the usual standard result (see, for instance, Ross, 1970). Q.E.D.

Our results thus far and Theorem 4.1.1 in Sandall (1974, p. 89) imply that for the finite horizon case an optimal sequence $(y_0, y_1, \ldots)$ is chosen as follows. We first determine a sequence $(\tilde{\delta}_0, \tilde{\delta}_1, \ldots)$, each element of which is a member of $\tilde{\Delta}$ and satisfies $\tilde{f}_{n-t} = \tilde{H}_{\tilde{\delta}_t} \tilde{f}_{n-t-1}$. An optimal strategy $(\delta_0, \delta_1, \ldots)$ is such that $\delta_t(x) = \{\tilde{\delta}_t[\Pi^k(x)]\}$ for all t. The decisions to be implemented, $(y_0, y_1, \ldots)$ are then chosen as $y_t(k) = \tilde{\delta}_t[\Pi^k(x)]$, where $x = x(t) = x(t-1)\,\mathscr{P}(y_{t-1})$. Thus, at time t when $z(t) = k$, the optimal decision is $y_t(k) \in C$. The theorem then states that $x(t) \ll x'(t)$ implies $y_t(k) \leqslant y_t'(k)$ for all $k \in Z$, where y_t (y_t') is associated with $x(t)$ $(x'(t))$.

In the theorem there is no indication as to how y_t might be related to the realization of the observation process at time t. The usual completely observed definition of a control-limit policy is the set of policies which are monotone with respect to the observation (which for the completely observed case is also the true state of the system). For the zero-memory problem considered here, this would correspond to controls being restricted to the set $D^* = \{y \in D : k \leqslant_Z k'$ implies $y(i, k) \leqslant y(i', k')\}$, where $\leqslant_Z$ is a given partial ordering on Z. Such a definition of control-limit policy is also identical to the usual definition when $S = Z$, $\leqslant_S = \leqslant_Z$, and $z(t) = s(t)$ for all t with probability one, i.e., $q_{jk} = 1$ when $j = k$. We formally state the set of all such control-limit policies as follows.

DEFINITION 3. $\Delta^* = \{\delta \in \Delta' : \delta$ maps $\tilde{\Delta}$ into $D^*\}$.

After the next preliminary result, a fourth assumption will be stated which together with the previously presented assumptions will guarantee that in selecting an optimal policy we need only consider those policies in the set Δ^*.

LEMMA. *Let $q = \{q_j\}$, $q' = \{q_j'\}$, and $x = \{x_j\}$ be arrays of identical size having nonnegative elements related by the partial $\leqslant'$. Assume $j \leqslant' k$ implies $q_j'/q_j' \leqslant' q_j/q_k$. Then for any subset K having the property that if $i \in K$ and $i \leqslant' i'$ then $i' \in K$,*

$$\sum_{j \in K} x_j q_j \Big/ \sum_j x_j q_j \leqslant \sum_{j \in K} x_j q_j' \Big/ \sum_j x_j q_j'.$$

Proof. Let K^c be the complement of K; define $A = \sum_{j \in K^c} x_j(q_j/q_m)$ and $B = \sum_{j \in K} x_j(q_j/q_m)$, where m is a minimal element of K. The left-hand side of the inequality in the problem statement then equals $B/(A + B)$; likewise, the right-hand side of the inequality equals $B'/(A' + B')$, where A' and B' are appropriately defined. Note that $B/(A + B) \leqslant B'(A' + B')$ is equivalent to $A'B \leqslant AB'$. The assumptions on q and q' then imply that $A' \leqslant A$ and $B \leqslant B'$, and the result is proved. Q.E.D.

We now state our fourth assumption.

ASSUMPTION A4. Q is such that if $k \leqslant_Z k'$ and $j \leqslant_S j'$, then $q_{jk'}/q_{j'k'} \leqslant q_{jk}/q_{j'k}$.

An important implication of (A4) is presented in the following result.

PROPOSITION. *Assume* (A4). *Then,* $k \leqslant_Z k'$ *implies that* $\Pi^k[x\mathscr{P}(y)] \prec \Pi^{k'}[x\mathscr{P}(y)]$ *for all* $y \in D$ *and* $x \in \Omega$.

Proof. Note that $\Pi_j{}^k[x\mathscr{P}(y)] = \tilde{x}_j q_{jk}/\sum_i \tilde{x}_i q_{ik}$, where

$$\tilde{x}_j = \sum_{i_1 \in S} \sum_{i_2 \in Z} x_{i_1 i_2} p_{i_1 j}[y(i_2)].$$

The result then follows from the definition of $\prec$ and the lemma. Q.E.D.

The proposition guarantees that assuming (A4) in addition to (A1), (A2), and (A3) implies that if $k \leqslant_Z k'$ then

$$y_t(k) = \tilde{\delta}_t[\Pi^k(x(t-1)\,\mathscr{P}(y_{t-1}))] \leqslant \tilde{\delta}_t[\Pi^{k'}(x(t-1)\,\mathscr{P}(y_{t-1}))] = y_t(k')$$

for $t \geqslant 1$. For the $t = 0$ case, note that $\Pi^k(x(0)) = \xi$ for all k and therefore $k \leqslant_Z k'$ implies $y_0(k) \leqslant y_0(k')$. We thus can restrict interest to elements in D^* for all t. These results are now stated in the following corollary.

COROLLARY. *Assume* (A1), (A2), (A3), *and* (A4). *Then, for the finite-horizon case, there exists an optimal strategy composed of policies that are members of* Δ^*. *For the discounted cost case* ($\beta < 1$), *there exists an optimal stationary strategy generated by a policy in* Δ^*.

We remark that (A4) is not a particularly restrictive assumption. For example, the two examples presented by White (1976c) which satisfied (A), (A2), and (A3) are easily shown to satisfy (A4) for properly chosen $\leqslant_Z$.

4. CONCLUSIONS

This paper has investigated the effect of information pattern and control problem description on the structural attributes of the optimal decision rule for a special Markov decision problem. Specifically, we have presented two generalized definitions of a control-limit policy for a replacement problem having a zero-memory, partially observed information pattern. For each definition, reasonable conditions were determined which were sufficient for the existence of optimal generalized control-limit policies. Investigation of the computational implications of these results in a subject of future research.

RECEIVED: October 1, 1976; REVISED: May 26, 1977

REFERENCES

ASTROM, K. J. (1965), Optimal control of Markov processes with incomplete state information, *J. Math. Analysis Appl.* **10**, 174.

DENARDO, E. (1967), Contraction mappings in the theory underlying dynamic programming, *SIAM Rev.* **9**, 165.

DERMAN, C. (1963), On optimal replacement rules when changes of state are Markovian, *in* "Mathematical Optimization Techniques" (R. Bellman, Ed.), pp. 201–210, Univ. of California Press, Berkeley.

KURTARAN, B.-Z. (1975), A concise derivation of the LQG one-step-delay sharing problem solution, *IEEE Trans. Automatic Control* **AC-20**, 808.

KUSHNER, H. (1975), "Introduction to Stochastic Control," Holt, Rinehart & Winston, New York.

PORTEUS, E. L. (1975), On the optimality of structured policies in countable stage decision processes, *Management Sci.* **22**, 148.

ROSS, S. M. (1970), "Applied Probability Models with Optimization Applications," Holden-Day, San Francisco.

SANDELL, N. R., JR. (1974), "Control of Finite-State, Finite-Memory Stochastic Systems," Sc.D. Thesis, Electrical Engineering Department, MIT, Cambridge, Mass.

SCARF, H. (1963), The optimality of (S, s) policies in the dynamic inventory problem, *in* "Mathematical Methods in the Social Sciences" (H. Scarf, D. Gilford, and M. Shelly, Eds.), Stanford Univ. Press, Stanford, Calif.

SMALLWOOD, R. D., AND SONDIK, E. J. (1973), The optimal control of partially observable Markov processes over a finite horizon, *Operations Res.* **21**, 1300.

SONDIK, E. J. (1971), "The Optimal Control of Partially Observable Markov Processes," Ph.D. Thesis, Stanford University, Stanford, Calif.

WHITE, C. C. (1975), Cost equality and inequality results for a partially observed stochastic optimization problem, *IEEE Trans. Syst. Man. Cybern.* **5**, 576.

WHITE, C. C. (1976a), Procedures for the solution of a finite-horizon, partially observed, semi-Markov optimization problem, *Operations Res.* **24**, 348.

WHITE, C. C. (1976b), Bounds on optimal cost for a partially observed machine subject to repair and Markov deterioration, to appear.

WHITE, C. C. (1976c), Optimal control-limit strategies for partially observed machine replacement, to appear.

WITSENHAUSEN, H. S. (1968), A counter-example in stochastic optimal control, *SIAM J. Control* **6**, 131.